hbase爬虫服务器部署

一、mac版

描述信息：爬虫程序，将数据保存到hbase中，从Mac电脑部署到服务器

服务器地址：172.17.106.106 用户名：root 密码：zx20150105

hosts:

10.0.1.22 cm

10.0.1.19 nn01

10.0.1.21 nn02

10.0.1.23 dn01

10.0.1.24 dn02

10.0.1.25 dn03

10.0.1.27 dn04

10.0.1.28 dn05

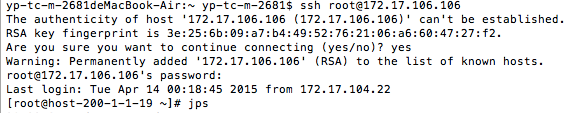
10.0.1.29 dn06

1.步骤一 mac连接服务器

在终端直接输入命令

ssh [root@172.17.106.106](mailto:root@172.17.106.106) root 用户名 172.17.106.106为服务器iP

如果想连接服务器的某个端口 ssh -p 8080 [root@172.17.106.106](mailto:root@172.17.106.106)

输入该命令后，如果服务器要输入密码，会提示你输入密码，输入就可以进行连接了

2.步骤二 配置服务器的hosts文件，这是hbase所需要的

在步骤一的基础上，输入命令 sudo vi /etc/hosts

之后再按键盘上的 I 键进入到文件的编辑模式 ，然后把

10.0.1.22 cm

10.0.1.19 nn01

10.0.1.21 nn02

10.0.1.23 dn01

10.0.1.24 dn02

10.0.1.25 dn03

10.0.1.27 dn04

10.0.1.28 dn05

10.0.1.29 dn06

复制到hosts文件中

然后按键盘上的ESC键，然后再输入wq!键对编辑后的文件进行保存，退出编辑模式

q!表示不保存

3.步骤三 将爬虫程序打成可运行的jar包，然后放到服务器上

将Mac上的文件传输到服务器

打开Mac的终端 （注：不是连接服务器后的终端），输入传输命令

scp /Users/yp-tc-m-2681/Documents/alibaba-shop.jar [root@172.17.106.106](mailto:root@172.17.106.106):~/soft

/Users/yp-tc-m-2681/Documents/alibaba-shop.jar：需要传输的文件路径（Mac）

root：服务器的用户名

172.17.106.106:服务器的ip

~/soft：将传输的文件放到服务器的某个目录下，这个地方放到soft目录下

~表示某个用户的目录 ，在Users目录下比如：Users/root 这时~就表示root目录下

4.步骤四 将爬虫程序在UNIX上运行起来

在服务器终端输入 java -jar XX.jar XX表示jar包的名称

注：当我们关闭终端的时候，刚启动的java进程也会关闭，因此需要让该进程与终端失去关联，即便关闭终端也不影响Java进程的运行。

所以在上面命令后，Java进程运行后，按键 ctrl+z然后输入bg 回车，再输入disown回车就可以了

即便是关闭终端也不影响Java程序的运行

注：如果Java进程里面有打印输入，此时有打印的时候，不能断断续续输入bg和disown命令，这两个命令必须一次性输完，，在没有打印输出的情况下。